STAT 611 Part 5
ARCH and GARCH Models

Zepu Zhang
May 4, 2011

1 ARCH(1)

Recall the ARMA model:

$$Y_t = (\phi_1 B + \phi_2 B^2 + \cdots + \phi_p B^p) Y_t + e_t - (\theta_1 B + \theta_2 B^2 + \cdots + \theta_q B^q) e_t$$

Our model for Y_{t+1} given history is that the only uncertain part is e_t, hence

$$Y_{t+1} \mid I_t \sim N\big((\phi_1 B + \phi_2 B^2 + \cdots + \phi_p B^p) Y_t - (\theta_1 B + \theta_2 B^2 + \cdots + \theta_q B^q) e_t,\ \sigma_e^2\big)$$

where I_t is called "information set"; it contains all information in the series up to time t, including past Y's and e's. Because e_t is an iid (normal) white noise, this ARMA model is a model for the conditional mean, whereas the conditional variance does not change—it's always σ_e^2.

Recall our ARMA simulations. The variability (level of fluctuation, or noisiness) in Y_t does not change. (Do not confuse the variability with change of the mean, or trends.)

However, we have seen, or can imagine, realistic series that contain periods of large variabilities, and periods of small variabilities. That is, the variance of Y tends to cluster, implying local correlations.

Such phenomena are studied in financial settings, and the variability is called "volatility" (think of stock market).

Note About the "return" example in the book: articles and textbooks on this topic typically uses examples about "return" (of investment, say), defined as

$$r_t = \frac{Y_t - Y_{t-1}}{Y_{t-1}}$$

An alternative definition is

$$r_t = \nabla \log Y_t = \log \frac{Y_t}{Y_{t-1}} = \log\left(1 + \frac{Y_t - Y_{t-1}}{Y_{t-1}}\right) \approx \frac{Y_t - Y_{t-1}}{Y_{t-1}}$$

The return, r_t, is considered to be a white noise (i.e. no serial correlation). But now we want to consider its "conditional variance". To be consistent with notation in the book, we'll use r_t in the sequel instead of e_t. (They mean the same thing.)

The idea is to build a model for the <u>conditional variance</u>. To simplify the matter, let's assume the <u>conditional mean</u> has been taken care of; the task is then to model the return r_t, endowing it with a variance that varies conditional on the magnitude of r_{t-1}. Let's assume

$$E(r_t \,|\, r_{t-1}, r_{t-2}, \dots) = 0$$
$$\mathrm{var}(r_t \,|\, r_{t-1}, r_{t-2}, \dots) = \omega + \alpha r_{t-1}^2$$

that is, the variance is a simple linear regression model of the squared level in the preceding time. To ensure $\mathrm{var}(r_t) > 0$, we require $\alpha \geq 0$. A more conventional way of writing this idea is

$$\begin{aligned} r_{t\,|\,t-1} &= \sigma_{t\,|\,t-1}\epsilon_t, \qquad \text{where } \epsilon_t \overset{\text{iid}}{\sim} N(0,1) \\ \sigma^2_{t\,|\,t-1} &= \omega + \alpha r_{t-1}^2 \end{aligned} \tag{1}$$

(Here ${}_{t\,|\,t-1}$ should be understood as conditional on all past times up to $t-1$, not just time $t-1$. In the above, of course, the model uses only the time $t-1$ of the history, but this does not have to be the case. As we'll see shortly, the model can be easily generalized to involve longer histories.)

Then, clearly,

$$r_{t\,|\,t-1} \sim N\big(0, \omega + \alpha r_{t-1}^2\big) \tag{2}$$

In contrast to ARMA, this ARCH(1) process has a constant conditional mean (always 0) but non-constant conditional variance. Therefore, it makes sense to construct ARMA models (for the conditional mean) with ARCH errors (for the conditional variance).

Note

$$E(r^2_{t\,|\,t-1}) = \mathrm{var}(r_{t\,|\,t-1}) = \omega + \alpha r_{t-1}^2$$

1.1 (Weak) Stationarity of the ARCH process

Let's examine the <u>unconditional</u> (or marginal) mean, variance, and auto-correlation of r_t:

$$E(r_t) = E_{r_{t-1}}\big(E(r_{t\,|\,t-1})\big) = E_{r_{t-1}}(0) = 0$$

$$\begin{aligned} \mathrm{var}(r_t) = E(r_t^2) = E_{r_{t-1}}\big(E(r^2_{t\,|\,t-1})\big) &= E_{r_{t-1}}(\omega + \alpha r_{t-1}^2) \\ &= E(\omega + \alpha r_{t-1}^2) \\ &= \omega + \alpha E(r_{t-1}^2) \\ &= \omega + \alpha\big(\omega + E(r_{t-2}^2)\big) \\ &= \cdots \\ &= \omega(1 + \alpha + \alpha^2 + \cdots) \\ &= \frac{\omega}{1-\alpha} \end{aligned}$$

This requires $0 \le \alpha < 1$. Note, although $\mathrm{var}(r_t)$ conditional on the history varies with the previous r_{t-1}, the marginal distribution of r_t has a constant variance.

$\mathrm{var}(r_t)$ is the marginal (or "unconditional") variance of the process, and can be denoted by σ^2.

For $h = 1, 2, \ldots,$

$$\mathrm{cov}(r_t, r_{t+h}) = E(r_t r_{t+h}) = E_{I_{t+h-1}}\big(E(r_t r_{t+h} \,|\, I_{t+h-1})\big) = E_{I_{t+h-1}}\big(r_t\, E(r_{t+h} \,|\, I_{t+h-1})\big) = 0$$

(using the fact the r_t is part of I_{t+h-1}, hence r_t is fixed conditional on I_{t+h-1}).

Observations:

1. The time invariance of $E(r_t)$, $\mathrm{var}(r_t)$, and $\mathrm{cov}(r_t, r_{t+h})$ establishes the weak stationarity of r_t.

2. $\mathrm{cov}(r_t, r_{t+h}) = 0$ establishes that r_t is a serially uncorrelated noise process (I think "uncorrelated" noise is called "white" noise), although it is not iid. Actually, it is serially dependent—clearly, the variance of r_t depends on r_{t-1}.

It turns out that $0 \le \alpha < 1$ is a sufficient and necessary condition for the (weak) stationarity of the ARCH(1) process.

1.2 AR representation of r_t^2

Adding

$$\begin{aligned} r_t^2 &= \sigma^2_{t\,|\,t-1}\epsilon_t^2 \\ \sigma^2_{t\,|\,t-1} &= \omega + \alpha r_{t-1}^2 \end{aligned}$$

we get

$$r_t^2 = \omega + \alpha r_{t-1}^2 + \sigma^2_{t\,|\,t-1}(\epsilon_t^2 - 1)$$

This is an AR(1) model for the squared process r_t^2, but with "innovations" (i.e. noises) $\sigma^2_{t\,|\,t-1}(\epsilon_t^2 - 1)$.

This hints at a way to check whether a conditional variance model should be considered: fit ARMA, get residuals $\hat{e}$, and plot the sample ACF and PACF of $\hat{e}^2$.

Let's define

$$\eta_t \equiv \sigma^2_{t\,|\,t-1}(\epsilon_t^2 - 1) = r_t^2 - \sigma^2_{t\,|\,t-1}$$

We can treat $\sigma^2_{t\,|\,t-1}$ as a prediction of r_t^2, then η_t is the "prediction error".

Properties of η_t: (1) $E(\eta_t) = 0$. (2) $\{\eta_t\}$ is serially uncorrelated. (3) η_t is uncorrelated with past squared returns. (To be verified: should it be "returns" or "squared returns" here?)

$\text{var}(r_t)$ can also be obtained from the AR(1) formulation of r_t^2 if we already assume $\{r_t\}$ is stationary. Taking expectation of both sides of the AR representation,

$$\text{var}(r_t) = \omega + \alpha\, \text{var}(r_t) + E(\eta_t) = \omega + \alpha\, \text{var}(r_t)$$

hence

$$\text{var}(r_t) = \frac{\omega}{1-\alpha}$$

1.3 Non-normality and fat tails of the ARCH process

We'll show that the marginal distribution of r_t is not normal. Specifically, it has heavier tails than normal. After some algebra we get

$$E(r_t^4) = \frac{3\omega^2}{(1-\alpha)^2}\frac{1-\alpha^2}{1-3\alpha^2}$$

For this to be finite, we need $\alpha^2 < 1/3$. In that case, the kurtosis of r_t is

$$\kappa = \frac{E(r_t^4)}{\left[E(r_t^2)\right]^2} = 3\frac{1-\alpha^2}{1-3\alpha^2} > 3 = \text{kurtosis of normal distribution}$$

(unless $\alpha = 0$), proving that r_t has "fat tails". If $\alpha^2 \geq 1/3$, $E(r_t^4)$ is infinite, the "fat tail" statement is still correct. The fat tail is a consequence of "volatility clustering".

This shows that to have finite moments in orders higher than 2, the coefficients need to be further restricted.

In summary, if $0 \leq \alpha < 1$, the process r_t is white noise and its unconditional distribution is symmetrically distributed around zero; this distribution has fat tails (compared to the normal distribution).

Why do we stress "non-normal" distribution here? Recall when we first introduced time series as a stochastic process, we said any group of the variable at a finite number of times form a multivariate random variable (or a random vector), and a somewhat simple situation is that we assume the joint distribution of normal. Here we have seen the process modeled by ARCH(1) can not be normal, even if the noise ϵ is normal. Normal is a typical distribution with the so-called "thin tail", which rarely gives rise to values that are extremely far from the mean. In comparison, the ARCH(1) process has a "heavier" (or "fatter") tail than a normal process.

1.4 Predicting conditional variances

$$\sigma^2_{t+1} = \omega + \alpha e_t^2 = (1-\alpha)\sigma^2 + \alpha e_t^2$$

This is a weighted average of the long-term (unconditional) variance σ^2 and the current squared process.

$$\begin{aligned}
\sigma^2_{t+h} \,|\, (r_t, r_{t-1}, \dots) &= E(r^2_{t+h}) \,|\, (r_t, r_{t-1}, \dots) \\
&= E(\sigma^2_{t+h\,|\,t+h-1}\epsilon^2_{t+h}) \,|\, (r_t, r_{t-1}, \dots) \\
&= E(\sigma^2_{t+h\,|\,t+h-1}) \,|\, (r_t, r_{t-1}, \dots) \\
&= \omega + \alpha E(r^2_{t+h-1}) \,|\, (r_t, r_{t-1}, \dots) \\
&= \omega + \alpha \sigma^2_{t+h-1} \,|\, (r_t, r_{t-1}, \dots)
\end{aligned}$$

provides a recursion.

Recursively using this relation, we get

$$\begin{aligned}
\sigma^2_{t+h} \,|\, (r_t, r_{t-1}, \dots) &= \omega + \alpha\big(\omega + \alpha\sigma^2_{t+h-2} \,|\, (r_t, r_{t-1}, \dots)\big) \\
&= \cdots \\
&= \omega + \omega\alpha + \cdots + \omega\alpha^{h-2} + \alpha^{h-1}(\omega + \alpha r_t^2) \\
&= \omega\frac{1-\alpha^h}{1-\alpha} + \alpha^h r_t^2 \\
&= (1-\alpha^h)\sigma^2 + \alpha^h r_t^2
\end{aligned}$$

We see the prediction of h-step-ahead variance is a weighted average of the long-term variance σ^2 and the squared return of the current time, r_t^2. The weight of the latter decreases exponentially with t.

1.5 MLE estimation

First, before estimation, we need to specify the order of the model. This is done through the AR model of r_t^2.

The key in writing down the likelihood is the conditional distribution

$$r_{t\,|\,t-1} \sim N(0, \omega + \alpha r^2_{t-1})$$

Note

$$L(\omega, \alpha; r_1, \dots, r_n) = f(r_1) f(r_2 \,|\, r_1) f(r_3 \,|\, r_2, r_1) \cdots f(r_n \,|\, r_{n-1}, \dots, r_1)$$

All the conditional terms are straightforward, whereas the first term, $f(r_1)$, naturally uses the marginal distribution

$$r_1 \sim N\Big(0, \frac{\omega}{1-\alpha}\Big)$$

2 ARCH(q)

ARCH(1) is generalized to ARCH(q) in obvious ways:

$$\sigma^2_{t\,|\,t-1} = \omega + \alpha_1 r^2_{t-1} + \alpha_2 r^2_{t-2} + \cdots + \alpha_q r^2_{t-q}$$

q is called the "ARCH order".

Properties of ARCH(q) can be discussed. For the most part, we can embed this discussion in the more general GARCH model below. But let's look at a few points.

Again define $\eta_t \equiv r_t^2 - \sigma^2_{t\,|\,t-1}$, we get an AR representation for r_t^2:

$$r_t^2 = \sigma^2_{t\,|\,t-1} + \eta_t = w + \alpha_1 r^2_{t-1} + \cdots + \alpha_q r^2_{t-q} + \eta_t$$

Similarly as in AR(1), we have $E(r_t) = 0$ and $E(\eta_t) = 0$.

Assuming stationarity, we take expectation on both sides of the AR form to get

$$\sigma^2 = \omega + \alpha_1 \sigma^2 + \cdots + \alpha_q \sigma^2$$

hence the unconditional variance is

$$\sigma^2 = \frac{\omega}{1 - \alpha_1 - \cdots - \alpha_q} \tag{3}$$

provided that $\alpha_1 + \cdots + \alpha_q < 1$.

Before estimating the model, how do we determine the order of the ARCH mode? Answer: it is the AR order of the r_t^2 process.

For MLE estimation, we use

$$\begin{aligned}
\sigma_1^2 &= \sigma^2 \\
\sigma^2_{2\,|\,1} &= \omega + \alpha_1 r_1^2 + (\alpha_2 + \cdots + \alpha_q)\sigma^2 \\
\sigma^2_{3\,|\,2} &= \omega + \alpha_1 r_2^2 + \alpha_2 r_1^2 + (\alpha_3 + \cdots + \alpha_q)\sigma^2 \\
\cdots &= \cdots \\
\sigma^2_{q\,|\,q-1} &= \omega + \alpha_1 r^2_{q-1} + \cdots + \alpha_{q-1} r_1^2 + \alpha_q \sigma^2 \\
\sigma^2_{t\,|\,t-1} &= \omega + \alpha_1 r^2_{t-1} + \cdots + \alpha_q r^2_{t-q}, \qquad t = q+1, \ldots, n
\end{aligned}$$

Note the marginal variance for r_1: we would have got the same result if we use

$$\sigma^2_{1\,|\,0} = \omega + \alpha_1 r_0^2 + \cdots \alpha_q r^2_{1-q}$$

and replace r_i^2, $i \le 0$, by its expectation σ^2, noticing the relation (3).

To predict a future variance conditional on $r_1, \dots, r_t$, note

$$\begin{aligned}
\sigma^2_{t+h\,|\,t} &= E(r^2_{t+h}) \,|\, (r_t, r_{t-1}, \dots) \\
&= E(\sigma^2_{t+h\,|\,t+h-1}\epsilon^2_{t+h}) \,|\, (r_t, r_{t-1}, \dots) \\
&= E(\sigma^2_{t+h\,|\,t+h-1}) \,|\, (r_t, r_{t-1}, \dots) \\
&= E(\omega + \alpha_1 r^2_{t+h-1} + \cdots + \alpha_q r^2_{t+h-q}) \,|\, (r_t, r_{t-1}, \dots) \\
&= \omega + \alpha_1 \sigma^2_{t+h-1\,|\,t} + \cdots + \alpha_q \sigma^2_{t+h-q\,|\,t}
\end{aligned}$$

gives a recursive formula, in which $\sigma^2_{s\,|\,t}$ is taken to be r^2_s when $1 \le s \le t$ and σ^2 when $s \le 0$.

In the derivation above,

$$\begin{aligned}
E(\sigma^2_{t+h\,|\,t+h-1}\epsilon^2_{t+h} \,|\, r_t, r_{t-1}, \dots) &= E(\sigma^2_{t+h\,|\,t+h-1} \,|\, r_t, r_{t-1}, \dots) E(\epsilon^2_{t+h} \,|\, r_t, r_{t-1}, \dots) \\
&= E(\sigma^2_{t+h\,|\,t+h-1} \,|\, r_t, r_{t-1}, \dots)
\end{aligned}$$

because of the independence between $\sigma^2_{t+h\,|\,t+h-1}$ and ϵ^2_{t+h}. A similar fact was used in the ARCH(1) section.

3 GARCH

The ARCH model is generalized to include auto-regression of σ^2_t:

$$\sigma^2_{t\,|\,t-1} = \omega + \beta_1 \sigma^2_{t-1\,|\,t-2} + \cdots + \beta_p \sigma^2_{t-p\,|\,t-p-1} + \alpha_1 r^2_{t-1} + \cdots + \alpha_q r^2_{t-q} \tag{4}$$

or

$$(1 - \beta_1 B - \cdots - \beta_p B^p)\sigma^2_{t\,|\,t-1} = \omega + (\alpha_1 B + \cdots + \alpha_q B^q) r^2_t \tag{5}$$

This is called GARCH(p, q) (with p GARCH orders and q ARCH orders).

Note: some books and software write ARCH order in front of GARCH order. Look carefully!

We assume all the coefficients, ω, α's, β's, are non-negative (which is not absolutely necessary for the model to be meaningful).

Note The GARCH generalization, i.e. inclusion of auto-regressive conditional variances, enhances the "memory" of the variance. For example, if a small return happens in a high volatility period, it will not immediately bring down the conditional variance by too much, thanks to the auto-regression in variance.

Unlike the ARMA model, which could be AR alone or MA alone, a GARCH model will always have the ARCH part (but may or may not have the 'G' generalization part), because we always want the connection between next-step variance and current r^2.

3.1 Example

Pages 290–294 shows exploratory plots for a simulated GARCH(1, 1) process.

Main points:

1. Plots 12.13 and 12.14, page 291, are sample ACF and PACF of $\{r_t\}$, demonstrating that r_t is serially uncorrelated.

2. Plots 12.15 and 12.16, page 292, are sample ACF and PACF of $\{|r_t|\}$, demonstrating that $|r_t|$ is serially correlated.

3. Plots 12.17 and 12.18, page 293, are sample ACF and PACF of $\{r_t^2\}$, demonstrating that r_t^2 is serially correlated.

3.2 ARMA representation for r_t^2

Again define $\eta_t = r_t^2 - \sigma^2_{t\,|\,t-1}$, then every σ^2 in (4) can be replaced by r^2 and η, leading to an ARMA(max$(p,q),p$) representation for r_t^2:

$$r_t^2 = \omega+(\beta_1+\alpha_1)r_{t-1}^2+\cdots+(\beta_{\max(p,q)}+\alpha_{\max(p,q)})r^2_{t-\max(p,q)}+\eta_t-\beta_1\eta_{t-1}-\cdots-\beta_p\eta_{t-p} \tag{6}$$

where $\beta_k = 0$ for all integers $k > p$ and $\alpha_k = 0$ for all integers $k > q$.

Therefore we can use techniques learned before to determine p and $\max(p,q)$ based on the $\{r_t^2\}$ series. If $\max(p,q) = p$, we can first fit a GARCH(p,p) model and then estimate q by examining the significance of the resulting ARCH coefficient estimates.

η_t has the same properties as that in the ARCH(1) model: (1) $E(\eta_t) = 0$. (2) $\{\eta_t\}$ is serially uncorrelated. (3) η_t is uncorrelated with past squared returns. (To be verified: should it be "returns" or "squared returns" here?)

3.3 Stationarity and fat tail

The GARCH process is (weakly) stationary if and only if

$$\sum_{i=1}^{p}\beta_i + \sum_{j=1}^{q}\alpha_j < 1$$

(under the assumption that $\beta_i \geq 0$ and $\alpha_j \geq 0$).

Take expectation on both sides of (6, we get the unconditional variance of the process $\{r_t\}$:

$$\sigma^2 = \frac{\omega}{1 - \sum_{i=1}^{q} \beta_i - \sum_{j=1}^{p} \alpha_j}$$

Note the analogy with the result for ARCH(1).

The stationary distribution of a GARCH model is generally fat-tailed.

3.4 Predicting conditional variances

Similar arguments give the following recursive formula:

$$\begin{aligned}
\sigma^2_{t+h\,|\,t} &= E(r^2_{t+h}\,|\,r_t, r_{t-1}, \dots) \\
&= E(\sigma^2_{t+h\,|\,t+h-1}\,|\,r_t, r_{t-1}, \dots) \\
&= \omega + \sum_{i=1}^{p} \beta_i \sigma^2_{t+h-i\,|\,t+h-i-1}\,|\,(r_t, r_{t-1}, \dots) + \sum_{j=1}^{q} \alpha_j \sigma^2_{t+h-j\,|\,t} \\
&= \omega + \sum_{i=1}^{p} \beta_i \sigma^2_{t+h-i\,|\,\min(t+h-i-1,\,t)} + \sum_{j=1}^{q} \alpha_j \sigma^2_{t+h-j\,|\,t}
\end{aligned}$$

where $\sigma^2_{u\,|\,v}$ is taken to be r_u^2 if $1 \leq u \leq v$ and σ^2 if $u \leq 0$. (If $u > v$, the values needs to be obtained by recursion.)

Alternatively, re-arranging (5) we get the MA(∞) form represented by

$$\sigma^2_{t\,|\,t-1} = \frac{w}{1 - \beta_1 B - \cdots - \beta_p B^p} + \frac{\alpha_1 B + \cdots + \alpha_q B^q}{1 - \beta_1 B - \cdots - \beta_p B^p} r_t^2$$

(to be continued...)

3.5 MLE estimation

First, we need to have an estimate of the orders. See page 294, below (12.3.4).

Like before, the likelihood is written based on the conditional distributions. The formulas are slightly simpler than those used for "predicting conditional variances" because we only need one-step-ahead predictions. The formulas are still recursive. Specifically,

$$\sigma^2_{t\,|\,t-1} = \omega + \sum_{i=1}^{p} \beta_i \sigma^2_{t-i\,|\,t-i-1} + \sum_{j=1}^{q} \alpha_j \sigma^2_{t-j\,|\,t-1}$$

where $\sigma^2_{u\,|\,v}$ is taken to be r_u^2 if $1 \le u \le v$ and σ^2 if $u \le 0$. (If $u > v$, the values needs to be obtained by recursion.)

Use the `R` package `tseries`.

4 Model diagnostics

Standardized residuals. (12.5.1), p 301.

Absolute value and square of residuals.

5 ARIMA (or even SARIMA) model with GARCH errors

The basic idea is to fit an ARIMA model first until the model seems adequate (i.e. no correlation left in residuals), then model the residuals by a GARCH model.